Voluntary versus Involuntary Perceptual Switching: Mechanistic Differences in Viewing an Ambiguous Figure

Michelle Umali (muu1@columbia.edu)
Center for Neurobiology & Behavior, Columbia University
1051 Riverside Drive, New York, NY, 10032, USA

Marc Pomplun (marc@cs.umb.edu)
Department of Computer Science, University of Massachusetts at Boston
100 Morrissey Blvd., Boston, MA 02125, USA

Abstract

Here we demonstrate the mechanistic differences between voluntary and involuntary switching of the perception of an ambiguous figure. In our experiment, participants viewed a 3D ambiguous figure, the Necker cube, and were asked to maintain one of two possible interpretations across four different conditions of varying cognitive load. These conditions differed in the instruction to freely view, make guided saccades, or fixate on a central cross. In the fourth condition, subjects were instructed to make guided saccades while unambiguous versions of the stimulus were intermittently flashed. Eye tracking measurements revealed consistent differences between voluntary and involuntary perceptual switching.

Introduction

In order to successfully interact with the environment, an individual must have up-to-date information with which to guide his or her behavior. However, an individual must be able to maintain that perception of the environment long enough to generate an appropriate response. Ambiguous figures, such as the Necker Cube, pose an interesting perceptual dilemma: more than one interpretation is possible without the stimulus changing. Viewing of the stimulus results in switching between these interpretations. In the case of the Necker cube (Figure 1), the object may be viewed as an upward facing cube (Figure 3a) or a downward facing cube (Figure 3b).

For nearly two centuries, a large amount of research has been devoted to uncovering the mechanisms underlying the maintenance of and switching between different perceptions of these ambiguous stimuli. Multiple theories have arisen to explain these phenomena and span the range of bottom-up to top-down mediated processes: switches result from neural fatigue within regions subserving a given percept (Toppino & Long, 1987), from guidance from a top-down drive to search (Leopold and Logothetis, 1999), or from top-down attempts to solve a perceptual question (Rock, 1975).

While earlier work focused on the measurement of switching rates, more recently the availability of precise eye-movement recording gave researchers a more detailed insight into the underlying perceptual mechanisms. Eye tracking is a powerful tool yielding multiple parameters such as gaze position, saccade amplitude, saccade frequency, blink frequency, and pupil size, which have been robustly correlated with cognitive function (see Rayner, 1998, for a review). Investigators utilizing this method have examined the regions within ambiguous figures that receive attention during a specific interpretation, as well as changes in eye movement parameters that may specify the time of switch.

For example, Ellis and Stark (1978) reported that prolonged fixation duration occurs at the time of perceptual switching. They also found regions within the Necker cube that when fixated, bias perception towards a particular interpretation of the cube. Pomplun, Ritter, and Velichkovsky (1996) had subjects press and hold a button A or B while perceiving interpretation A or B, respectively, of various ambiguous images. They separated the obtained eye-fixation positions into two groups, according to the button that was being pressed during the fixation. For most images it was found that subjects looked at distinct regions in the pictures for interpretations A and B. Interestingly, the distance of these two regions was maximal if the fixations were not separated at the time of button presses, but at about 900 to 1000 ms before each button press. This finding suggests that changes in low-level visual mechanisms precede conscious perceptual switches by a substantial amount of time. However, for line drawings such as the Necker cube in which each line is equally important across interpretations, no difference in eye movements between the two interpretations was found.

More recent work by Ito et al. (2003) identified the eye-movement variables blink frequency, saccade frequency, and saccade direction as sensitive to perceptual switching. Moreover, the study attributed the finding by Ellis and Stark (1978) of prolonged fixations during switches to biases arising from an inadequate sampling frequency. Instead of describing a single time point that defines the perceptual switch, Ito et al. depict a process that extends over the range of seconds. While changes in saccade direction were specific to their stimulus, their finding of peaks in blink frequency and saccade frequency occurring before and after button pressing (to indicate the switch) gave new insight into the underlying perceptual mechanisms.

However, one of the most fundamental and fascinating aspects of viewing ambiguous images, namely subjects' intentional control, has not been systematically studied with these fine-grained methods. In order to comply with instructions favoring switching or maintenance, individuals can modulate their rates of perceptual switching (Toppino, 2003).

Therefore, the goal of the current study was to expand upon previous work by examining the differences between voluntary and involuntary perceptual switches between interpretations of the Necker cube. If subjects are instructed to maintain a given perspective, involuntarily switching to the opposite view should require less cognitive effort than making a voluntary switch back to the favored perspective. This relationship should also be sustained across tasks of increasing cognitive demand. We will use eye tracking in order to obtain physiological responses reflecting cognitive performance and to analyze their robustness across experimental conditions. Specifically we will examine the time course of blink frequency, saccade frequency, and saccade amplitude across these two types of perceptual switches.

Method

Participants

Ten students (3F; aged 21-35), from UMass Boston participated in the experiment. They were paid for their participation and did not have any information about the nature of the study.

Apparatus

Eye movements were recorded with the SR Research Ltd. EyeLink-II system, which operates at a sampling rate of 500 Hz and measures a subject's gaze position with an average error of 0.5 degrees of visual angle. Stimuli were presented on a 21-inch Dell Trinitron monitor with a refresh rate of 85 Hz and a screen resolution of 1280 by 1024 pixels.

Materials

Figure 1 shows the Necker cube stimulus as seen by the subjects across all four conditions. Two interpretations, a downward facing cube and an upward facing cube, are possible. For two experimental conditions subjects were asked to make saccades switching back and forth between two specific vertices on the cube in time with a presented tone. The arrow in Figure 2 shows the path of the saccades the subjects were instructed to make. In another experimental condition, unambiguous versions of the Necker cube, which depicted the interpretation opposite to the one that participant was to hold, were flashed every 5 s (Figure 3).

Procedure

Before data was collected, participants received 30 minutes of instruction regarding the tasks, orientation with the response pad, and practice perceiving the two interpretations of the Necker cube.

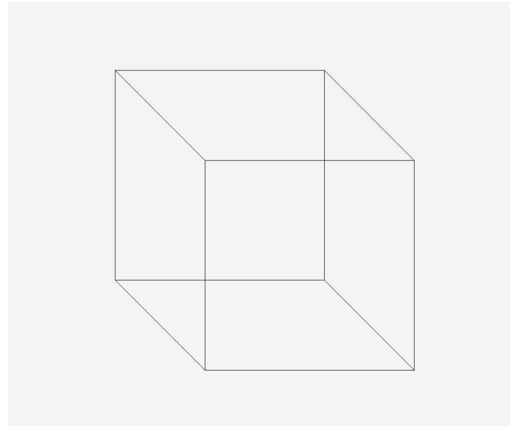

Figure 1: Necker cube stimulus as seen by the participant.

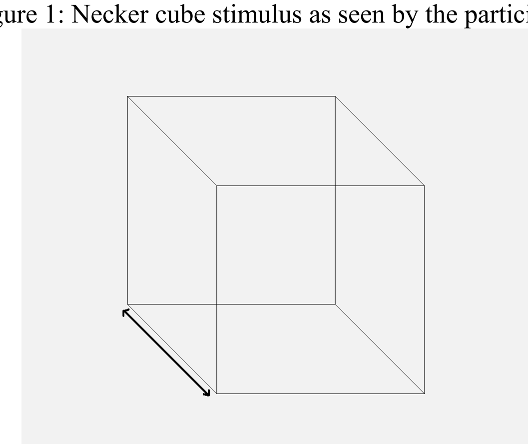

Figure 2: Instructed saccade targets. In experimental conditions 3 and 4, subjects were asked to maintain a perception while making saccades to the vertices indicated when they heard a tone.

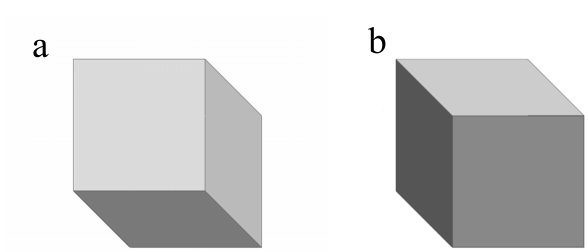

Figure 3: The two unambiguous versions of the stimulus showing the two possible interpretations of the Necker cube are presented here (upward = a, downward = b). In experimental condition 2, unambiguous stimuli were flashed in place of the original Necker cube every 5 s.

The experimental design consisted of four conditions: 1) the participant was instructed to view the stimulus and look wherever he or she wishes ("Free Viewing"), 2) the participant was instructed to fixate on a marker in the center of the cube ("Fixation"), 3) the participant was instructed to make alternating saccades to the two bottom left vertices in time with a tone presented every 500 ms ("500ms Saccades"), and 4) the instructions to the participant were the same as in 3, but an unambiguous version of the stimulus was flashed in place of the Necker cube every 5 seconds for a duration of 150 ms, coincident with the tone presented every 500ms ("Flashing"). For constancy across conditions, the tone was present in every condition, but participants were instructed to ignore the tone during the Free Viewing and Central Fixation trials.

The order of the conditions was randomized, and between conditions participants received brief reminders as to instructions for the upcoming task. For each condition there were two 45-second trials, one in which they were to hold the upward perspective and the other where they were to hold the downward perspective. Half the subjects were to hold the upward perspective in the first trial and the downward perspective on the second trial. The reverse was true for the other half of the subjects. Subjects indicated which interpretation they were holding by pressing a corresponding button on a game pad. Before each trial, the eye tracker was re-calibrated using a single central fixation target to sustain high precision of measurement.

Data Analysis

Similar to Ito et al. (2003), a moving window technique was used to analyze eye movement variables as functions of time relative to button presses or flashes. To compute the value of a variable at a certain relative time t, its average value across all intervals from t – 1000 ms to t + 1000 ms relative to the relevant event was determined. To derive the graphs in the present paper, the center of this window was moved in steps of 20 ms from -5000 ms to 5000 ms relative to a button press, and from 0 to 5000 ms relative to a flash in the Flashing condition. Blink frequency was measured as the number of blinks per second, that is, the average number of blinks found in the 2-second window, divided by 2. Saccade frequency was measured analogously, and saccade amplitude was determined as the average amplitude of all saccades that were measured within a 2-second window.

For the purpose of statistical analysis, values of the same variable were split into ten different bins according to their time of measurement relative to button presses. The first bin contained the data for the interval -5000 ms to – 4001 ms, the second one the data from -4000 to -3001 ms, and so on, and the tenth bin contained the data from 4000 to 4999 ms. The three-way repeated measures analyses of variance (ANOVAs) computed on the eye-tracking variables had the factors "experimental condition" (4 levels: Free Viewing, Fixation, 500ms Saccades, and Flashing), "type of switch" (2 levels: voluntary and involuntary), and "relative time" (10 levels, referring to the ten relative time intervals described above). For the analysis of variables relative to flashes, the ten time intervals were chosen to span the interval from 0 ms to 5000 ms relative to the flashes.

Results and Discussion

Mean duration of the interval between perceptual switches (i.e. the mean holding duration of the voluntary percept or involuntary percept) was determined for all four experimental conditions (Figure 4). A two-way ANOVA with factors "experimental condition" and "percept" (voluntary vs. involuntary) revealed that, across all conditions, the duration of perceptual switching intervals was longer for voluntary than involuntary perception, $F(1; 9) = 15.89$, $p < 0.005$. The holding durations for the involuntary percepts were similar across all conditions, and so were the durations for the voluntary percepts, except for the Flashing condition. In that experimental condition, the mean duration for voluntary switches was significantly shorter compared with the other conditions. However, due to substantial individual differences across subjects, neither the main effect of experimental condition nor the interaction of the two factors reached significance, both $Fs(3;27) < 2.04$, $ps > 0.1$. This result demonstrates that subjects were able to intentionally bias their perception in favor of the instructed type of interpretation of the Necker cube.

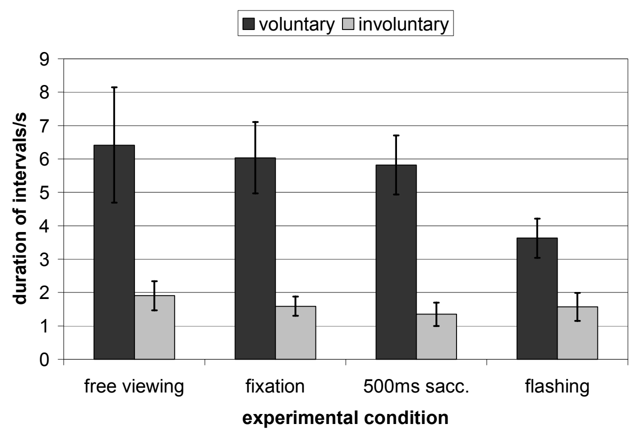

Figure 4: Mean duration of perceptual intervals across conditions. Values with standard error bars are shown for voluntary (dark bars) and involuntary switches (light bars).

The close examination of eye movement parameters across the two switch types, however, promised to be more insightful than interval durations. Data were aligned relative to the button press in order to investigate how these parameters were modulated around the time of reporting of a switch. Blink frequency, saccade frequency, and saccade amplitude were determined for each of the four conditions and separated for voluntary and involuntary switches.

The three-way ANOVA for blink frequency revealed a main effect of relative time, $F(9; 81) = 4.19$, $p < 0.001$, and a significant interaction between the factors relative time and type of switch, $F(9; 81) = 4.34$, $p < 0.001$; no other effects were found. These results indicate that blink frequency is sensitive to perceptual switches, and that its time course differs between voluntary and involuntary switches. As can be seen in Figure 5, first column, blink frequency increased around the time of button pressing across all conditions. Interestingly, the blink frequency peak for the voluntary switches always preceded those for involuntary switches, explaining the significant interaction between the factors relative time and type of switch.

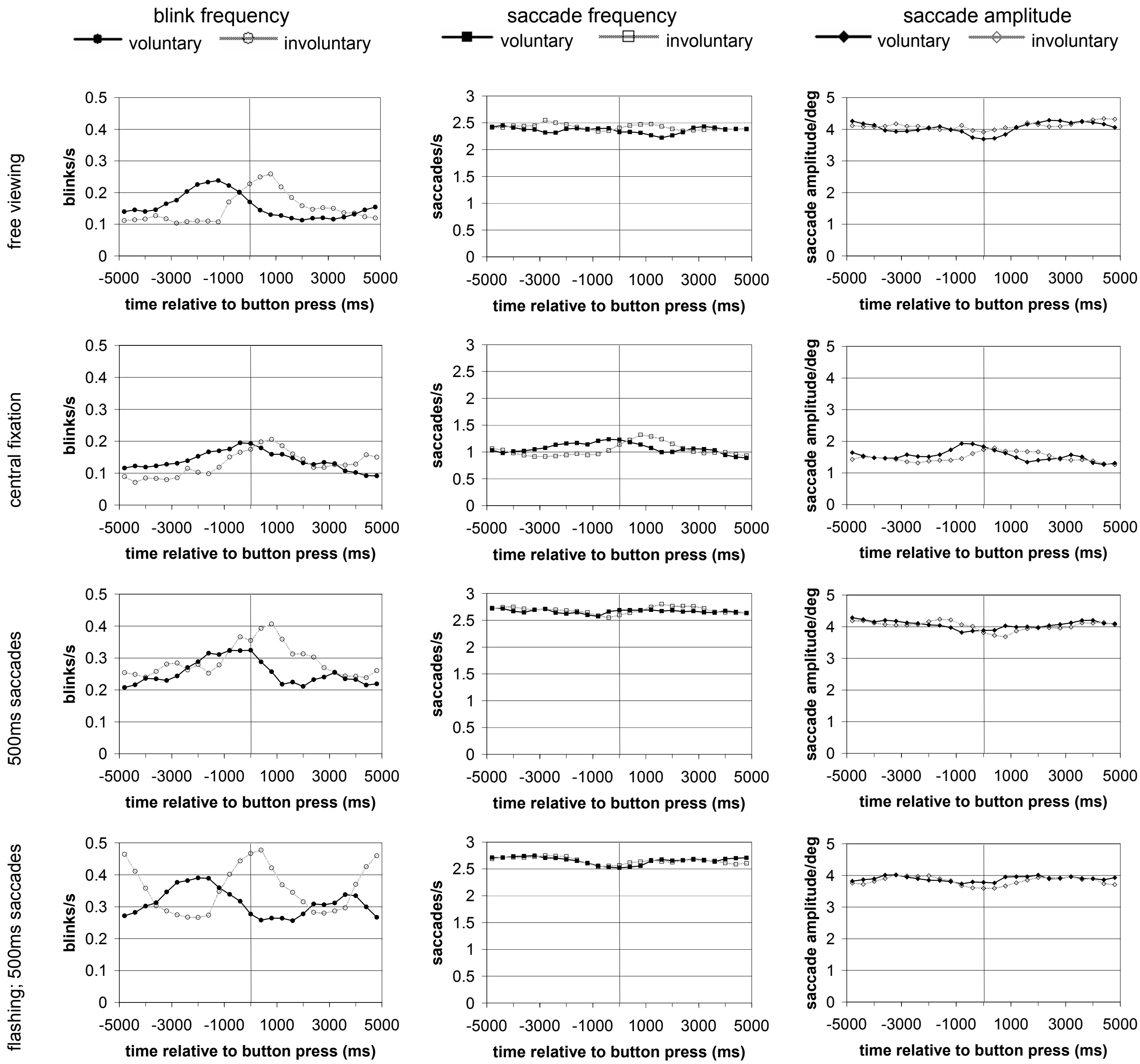

Figure 5. Time course of blink frequency (column 1), saccade frequency (column 2), and saccade amplitude (column 3) across all conditions. Data occurring during voluntary switches (filled symbols) and involuntary switches (open symbols) are shown relative to the time of the respective button press.

Although the factor experimental condition did not reach significance, the diagrams suggest subtle differences across conditions. During the Free Viewing condition, the effects of voluntary and involuntary switches were of similar strength, but with a large time difference between their peaks (1000 ms before and after the button press, respectively). For the Fixation condition, blink frequency showed a rather similar time course between voluntary and involuntary switches. The peaks in blink frequency for both voluntary and involuntary switches occurred near the time of button presses and were of similar magnitude. During the 500ms Saccades and Flashing conditions – the two saccadic conditions - blink frequency was predominantly greater in magnitude for involuntary switches than for voluntary switches, and the first voluntary peak always clearly preceded the elevation for involuntary switching. The Flashing condition is unique in that it involves a periodic stimulus manipulation imposing a systematic temporal influence on the psychophysical data. Therefore, this condition will receive a separate analysis below.

All in all, the blink frequency data reveal a robust divergence in time courses for voluntary and involuntary switching, and they also suggest a divergence in magnitude that varies across conditions. Blink frequency is known to be inversely correlated with cognitive effort (e.g., Veltman and Gaillard, 1998). In the present context, however, it is unclear whether changes in blink frequency actually indicate varying cognitive effort or rather the presence of other mechanisms underlying perceptual switching. It is hard to explain why cognitive effort should decrease around the time of perceptual switches and button responses. At any rate, the pattern of results suggests that voluntary and involuntary switches involve different activation sequences of underlying neural mechanisms.

This significant divergence in time course between voluntary and involuntary switches carried over to saccade frequency and amplitude parameters, though to a much lesser extent. The three-way ANOVAs for these two variables showed a significant main effect for the factor experimental condition, both Fs(3; 24) > 30.06, ps < 0.001. Moreover, for both variables the two-way interaction of experimental condition and relative time, both Fs(27; 216) > 1.92, ps < 0.01, and the three-way interaction of experimental condition, relative time, and type of switch, both Fs(27; 243) > 1.79, ps < 0.05, were significant.

The main effect of the experimental condition is an obvious consequence of the different instructions with regard to subjects' saccadic behavior. For example, Figure 5 shows that both saccade frequency (second column) and amplitude (third column) were smaller in the Fixation condition than in all other conditions. The significant two-way and three-way interactions demonstrate for both saccade variables that their temporal behavior varied across combinations of experimental condition and type of switch. Generally fewer and shorter saccades occurred during switches, but the opposite situation occurred during the Fixation condition. This result suggests that perceptual switching disrupts task performance and makes subjects more likely to deviate from their assigned task. Taken together, the findings with regard to saccade frequency and amplitude yield further evidence for a differential involvement of neural mechanisms in voluntary and involuntary switches – voluntary switches have an earlier effect on these variables than involuntary switches.

Finally, the Flashing condition needed to be analyzed in a separate way. This condition was distinct from the others in that there was a 5-second rhythm of the flashing of an unambiguous cube, which represented a percept opposite to the one participants were trying to maintain. We were interested in examining how eye movement parameters and button press times varied in temporal relation to these flashes (Figure 6).

One-way ANOVAs with the factor relative time showed no significant effect on saccade amplitude or saccade frequency, both Fs(9; 81) < 1. Blink frequency, however, was significantly influenced by the relative time, F(9; 81) = 7.33, p < 0.001. As can be seen in Figure 6, blink frequency exhibited a strong modulation with a peak centered at 1000 ms following the flash, after which values steadily declined to a minimum at 4200 ms. The following increase in frequency was due to our method of using a 2-second window for data analysis. It does not indicate subjects' anticipation of the next flash, but an effect of temporal averaging within the 2-second window. Through this averaging, changes in variables that occur, for example, 200 ms after a flash still influence the data from 800 ms before a flash (= 4200 ms after a flash) to 1200 ms after a flash.

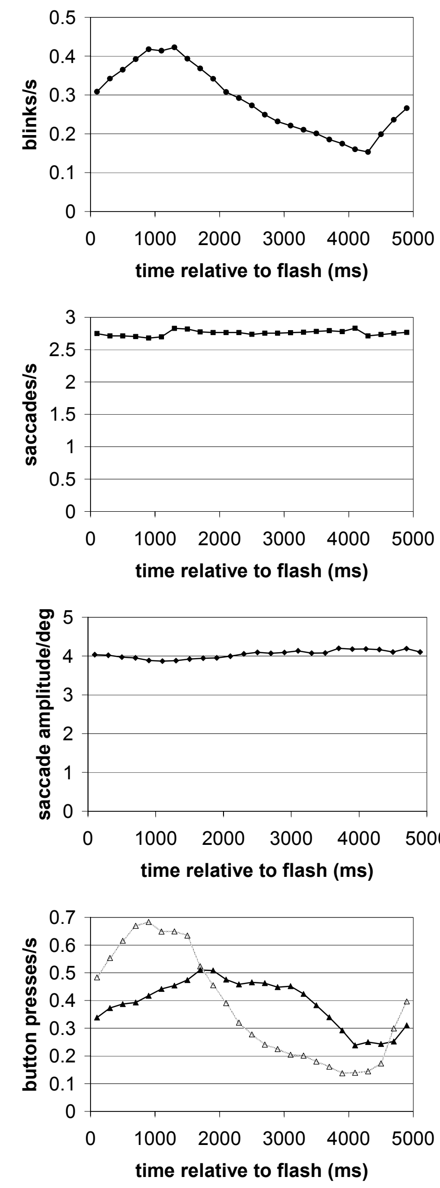

Figure 6. Time course of eye movement parameters and response frequency relative to the time of the flash of the unambiguous stimulus. From top to bottom: Blink frequency, saccade frequency, and saccade amplitude data represent pooled results from voluntary and involuntary switching. Data for response frequency is divided between voluntary (filled triangles) and involuntary (open triangles) perceptual switches.

The time points of blink frequency maxima and minima were well aligned with those for response frequency of involuntary switches. The frequency of involuntary perceptual reversals was highest approximately 1000 ms after the unambiguous figure was flashed, and it was lowest at around 4400 ms after a flash (= 600 ms before a flash). For voluntary reversals, the peak occurred later and was less pronounced. A two-way ANOVA with factors relative time and type of switch revealed a significant main effect of relative time on response frequency, $F(9; 81) = 2.71$, $p < 0.05$, and a significant interaction between the two factors, $F(9; 81) = 3.11$, $p < 0.01$.

Taken together, the analysis of psychophysical data relative to flashing the “unwanted” interpretation of the cube demonstrated a strong bias of subjects’ perception towards that interpretation. One second after a flash, subjects are more than four times as likely to switch to the involuntary percept as they are four seconds after a flash. This effect of a brief 150 ms flash is unexpectedly strong. It is also interesting to see that the average time from a flash to the manual report of the corresponding percept is about 1000 ms. This corresponds to the delay between the start of a perceptual switch – as indicated by eye-movement parameters – and its manual report found by Pomplun and colleagues (1996). A possible interpretation of this correspondence is that, once triggered, the sequence of neural mechanisms underlying perceptual switches takes about 1000 ms to complete, minus some manual reaction time. This seems to work in similar ways for externally and internally triggered switches.

In summary, our most important finding is the evidence for two distinct types of perceptual switching that occur during viewing of an ambiguous figure. These switches exhibit different time courses of underlying mechanisms from one another with effects on eye-movement variables occurring earlier during voluntary switches relative to involuntary switches. The finding by Pomplun et al. (1996) of changes in eye-movement variables preceding button responses by 900-1000 ms was replicated in our study for involuntary switches but not voluntary ones. One possibility for this difference may be that the cognitive mechanisms underlying perceptual switching are recruited actively and thus more quickly during voluntary switches.

References

Ellis S.R. and Stark L (1978). Eye movements during viewing of Necker cubes. *Perception*, 7, 575-581.

Ito J., Nikolaev A.R., Luman M., Aukes M.F., Nakatani C., van Leeuwen C. (2003). Perceptual switching, eye movements, and the bus paradox. *Perception*, 32, 681-698.

Leopold D.A. & Logothetis N.K. (1999). Multistable phenomena: changing views in perception. *Trends in Cognitive Sciences,* 3, 254-264.

Pomplun M, Ritter H, & Velichkovsky B. (1996). Disambiguating complex visual information: Towards communication of personal views of a scene. *Perception*, 25(8), 931-48.

Rayner K. (1998) Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422.

Rock I. (1975). *An Introduction to perception.* New York: Macmillan.

Scotto M.A. (1991). Smooth periodic eye movements can entrain perceptual alternation. *Perceptual and Motor Skills,* 73, 835-843.

Scotto M.A., Oliva G.A., Tuccio M.T. (1990). Eye movements and reversal rates of ambiguous patterns. *Perceptual and Motor Skills*, 70, 1059-1073.

Toppino T.C. (2003). Reversible-figure perception: mechanisms of intentional control. *Perception & Psychophysics,* 65(8), 1285-1295.

Toppino T.C. & Long G.M. (1987). Selective adaptation with reversible figures: Don't change that channel. *Perception & Psychophysics*, 42, 37-48.